```
        10         20         30         40         50
GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATC
        60         70         80         90        100
TGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTT
       110        120        130        140        150
GGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAG
       160        170        180        190        200
GCTTGACCGACAATTGAGCTCGGTACCCGGGGAGATCCGGTAAGGACCAG
       210        220        230        240        250
CTTCTTTGGGAGAGAACAGACGCAGGGGCGGGAGGGAAAAAGGGAGAGGC
       260        270        280        290        300
AGACGTCACTTCCCCTTGGCGGCTCTGGCAGCAGATTGGTCGGTTGAGTG
       310        320        330        340        350
GCAGAAAGGCAGACGGGGACTGGGCAAGGCACTGTCGGTGACATCACGGA
       360        370        380        390        400
CAGGGCGACTTCTATGTAGATGAGGCAGCGCAGAGGCTGCTGCTTCGCCA
       410        420        430        440        450
CTTGCTGCTTCACCACGAAGGAGTTCCCGTGCCCTGGGAGCGGGTTCAGG
       460        470        480        490        500
ACCGCTGATCGGAAGTGAGAATCCCAGCTGTGTGTCAGGGCTGGAAAGGG
       510        520        530        540        550
CTCGGGAGTGCGCGGGGCAAGTGACCGTGTGTGTAAAGAGTGAGGCGTAT
       560        570        580        590        600
GAGGCTGTGTCGGGGCAGAGGCCCAAGATCTCAAGGGCCCATAACATGTG
       610        620        630        640        650
TACCATCGATTGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAG
       660        670        680        690        700
GGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCC
       710        720        730        740        750
CAAAGCTTGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTT
       760        770        780        790        800
CGCGCTTTCCCCTGACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAC
       810        820        830        840        850
CGGATCCTCTAGAGTCGACCTGCAGGCATGCAGAAGACAATTAGCAGGCA
       860        870        880        890        900
TGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCT
       910        920        930        940        950
GGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCG
```

Fig. 1A

```
       960        970         980       990       1000
GCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCT
      1010       1020       1030       1040       1050
AGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCG
      1060       1070       1080       1090       1100
GCTTTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTT
      1110       1120       1130       1140       1150
AGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTC
      1160       1170       1180       1190       1200
ACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGG
      1210       1220       1230       1240       1250
AGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTC
      1260       1270       1280       1290       1300
AACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTC
      1310       1320       1330       1340       1350
GGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATT
      1360       1370       1380       1390       1400
AATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCC
      1410       1420       1430       1440       1450
CAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAG
      1460       1470       1480       1490       1500
GTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGC
      1510       1520       1530       1540       1550
ATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCG
      1560       1570       1580       1590       1600
CCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAAT
      1610       1620       1630       1640       1650
TTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCC
      1660       1670       1680       1690       1700
AGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTC
      1710       1720       1730       1740       1750
CCGGGAGCTTGTATATCCATTTTCGGATCTGATCAGCACGTGTTGACAAT
      1760       1770       1780       1790       1800
TAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAA
      1810       1820       1830       1840       1850
CTAAACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCG
      1860       1870       1880       1890       1900
ACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGG
```

Fig. 1B

```
      1910       1920       1930       1940       1950
GACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCT
      1960       1970       1980       1990       2000
GTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCT
      2010       2020       2030       2040       2050
GGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTC
      2060       2070       2080       2090       2100
GTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGG
      2110       2120       2130       2140       2150
CGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACT
      2160       2170       2180       2190       2200
GCGTGCACTTCGTGGCCGAGGAGCAGGACTGACACGTGCTACGAGATTTC
      2210       2220       2230       2240       2250
GATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCG
      2260       2270       2280       2290       2300
GGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCT
      2310       2320       2330       2340       2350
TCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGC
      2360       2370       2380       2390       2400
AATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAG
      2410       2420       2430       2440       2450
TTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGT
      2460       2470       2480       2490       2500
CGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTG
      2510       2520       2530       2540       2550
TGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCAT
      2560       2570       2580       2590       2600
AAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG
      2610       2620       2630       2640       2650
CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTG
      2660       2670       2680       2690       2700
CATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCG
      2710       2720       2730       2740       2750
CTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGC
      2760       2770       2780       2790       2800
GGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAA
      2810       2820       2830       2840       2850
TCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
```

Fig. 1C

```
      2860       2870       2880       2890       2900
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCC
      2910       2920       2930       2940       2950
CCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC
      2960       2970       2980       2990       3000
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGT
      3010       3020       3030       3040       3050
GCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTC
      3060       3070       3080       3090       3100
TCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTC
      3110       3120       3130       3140       3150
AGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCC
      3160       3170       3180       3190       3200
CGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
      3210       3220       3230       3240       3250
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGG
      3260       3270       3280       3290       3300
ATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTG
      3310       3320       3330       3340       3350
GCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGC
      3360       3370       3380       3390       3400
TGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAA
      3410       3420       3430       3440       3450
CAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTAC
      3460       3470       3480       3490       3500
GCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
      3510       3520       3530       3540       3550
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGA
      3560       3570       3580       3590       3600
TTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTT
      3610       3620       3630       3640       3650
TAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
      3660       3670       3680       3690       3700
GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCC
      3710       3720       3730       3740       3750
ATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTT
      3760       3770       3780       3790       3800
ACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGG
```

Fig. 1D

```
        3810      3820      3830      3840      3850
CTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGA
        3860      3870      3880      3890      3900
AGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCG
        3910      3920      3930      3940      3950
GGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTG
        3960      3970      3980      3990      4000
CCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCA
        4010      4020      4030      4040      4050
TTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTT
        4060      4070      4080      4090      4100
GTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTA
        4110      4120      4130      4140      4150
AGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCT
        4160      4170      4180      4190      4200
CTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTC
        4210      4220      4230      4240      4250
AACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
        4260      4270      4280      4290      4300
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTG
        4310      4320      4330      4340      4350
CTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACC
        4360      4370      4380      4390      4400
GCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTT
        4410      4420      4430      4440      4450
CAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGG
        4460      4470      4480      4490      4500
CAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACT
        4510      4520      4530      4540      4550
CATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTC
        4560      4570      4580      4590      4600
TCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGG
        4610      4620      4630
GTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC
```

Fig. 1E

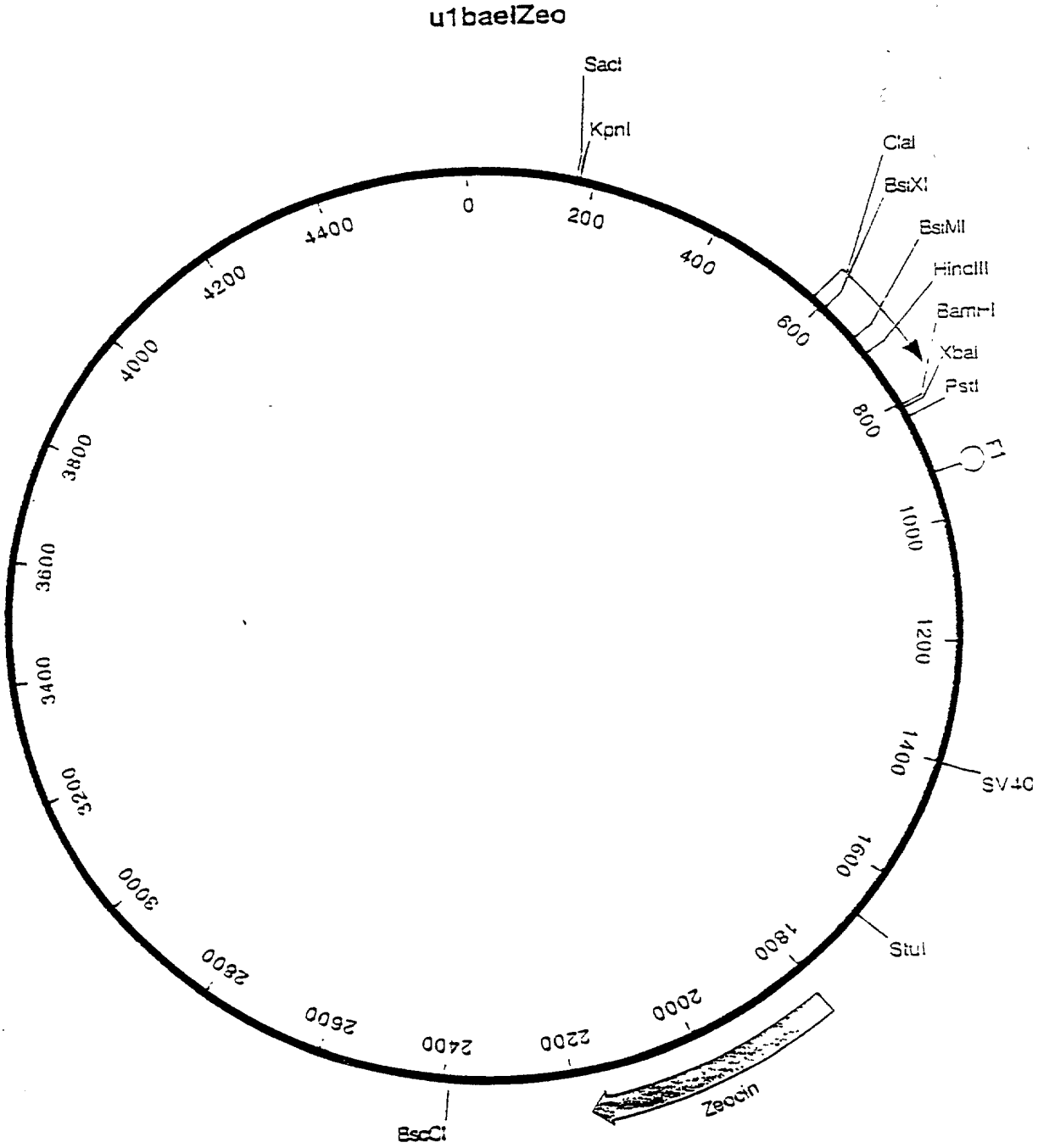
u1baeIZeo
SacI
KpnI
ClaI
BsiXI
BsiMI
HindIII
BamHI
XbaI
PstI
F1
SV40
StuI
Zeocin
BscCI
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
3200
3400
3600
3800
4000
4200
4400

Fig. 2

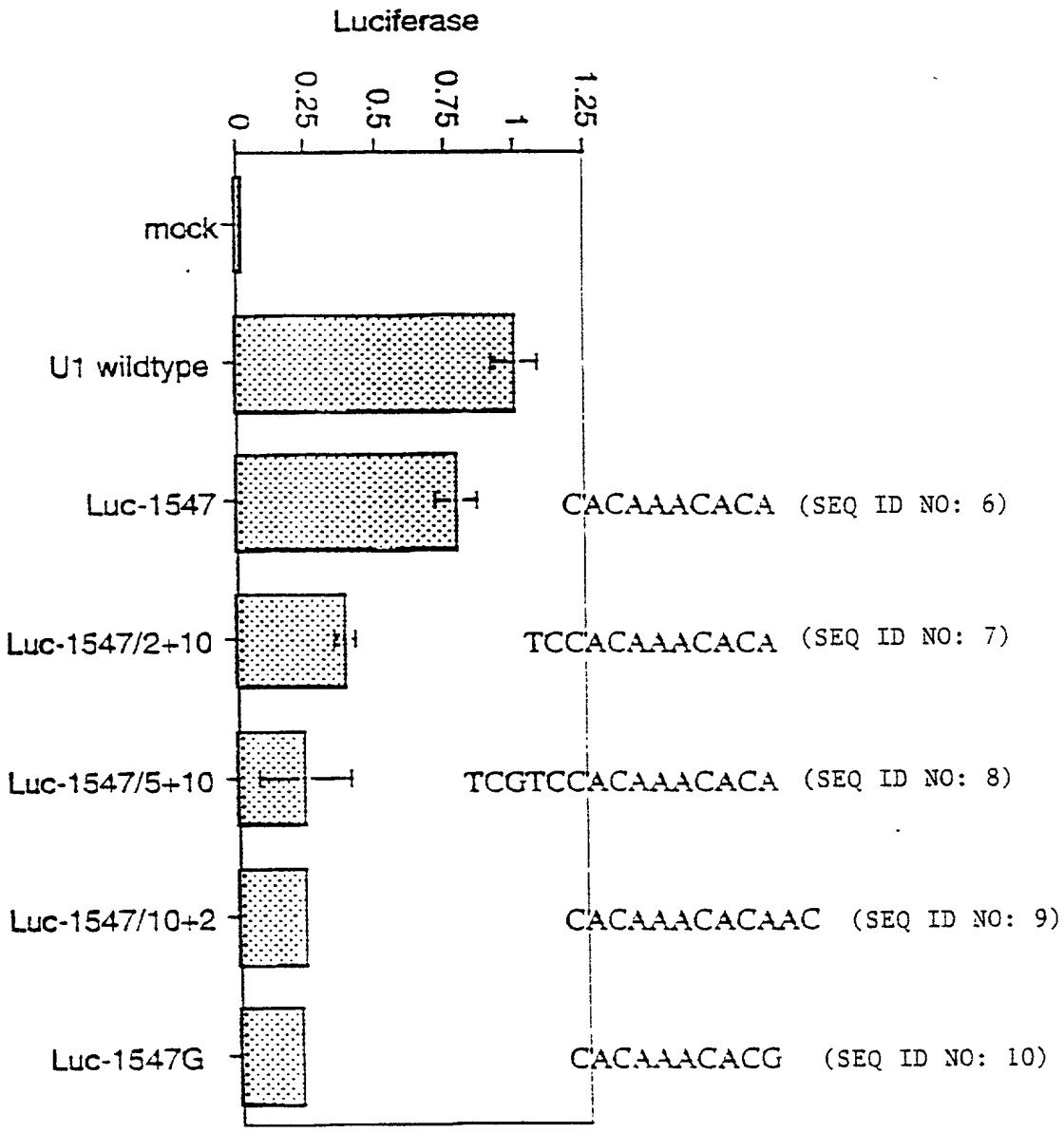
Luciferase
0
0.25
0.5
0.75
1
1.25
mock
U1 wildtype
Luc-1547
CACAAACACA (SEQ ID NO: 6)
Luc-1547/2+10
TCCACAAACACA (SEQ ID NO: 7)
Luc-1547/5+10
TCGTCCACAAACACA (SEQ ID NO: 8)
Luc-1547/10+2
CACAAACACAAC (SEQ ID NO: 9)
Luc-1547G
CACAAACACG (SEQ ID NO: 10)

Fig. 3

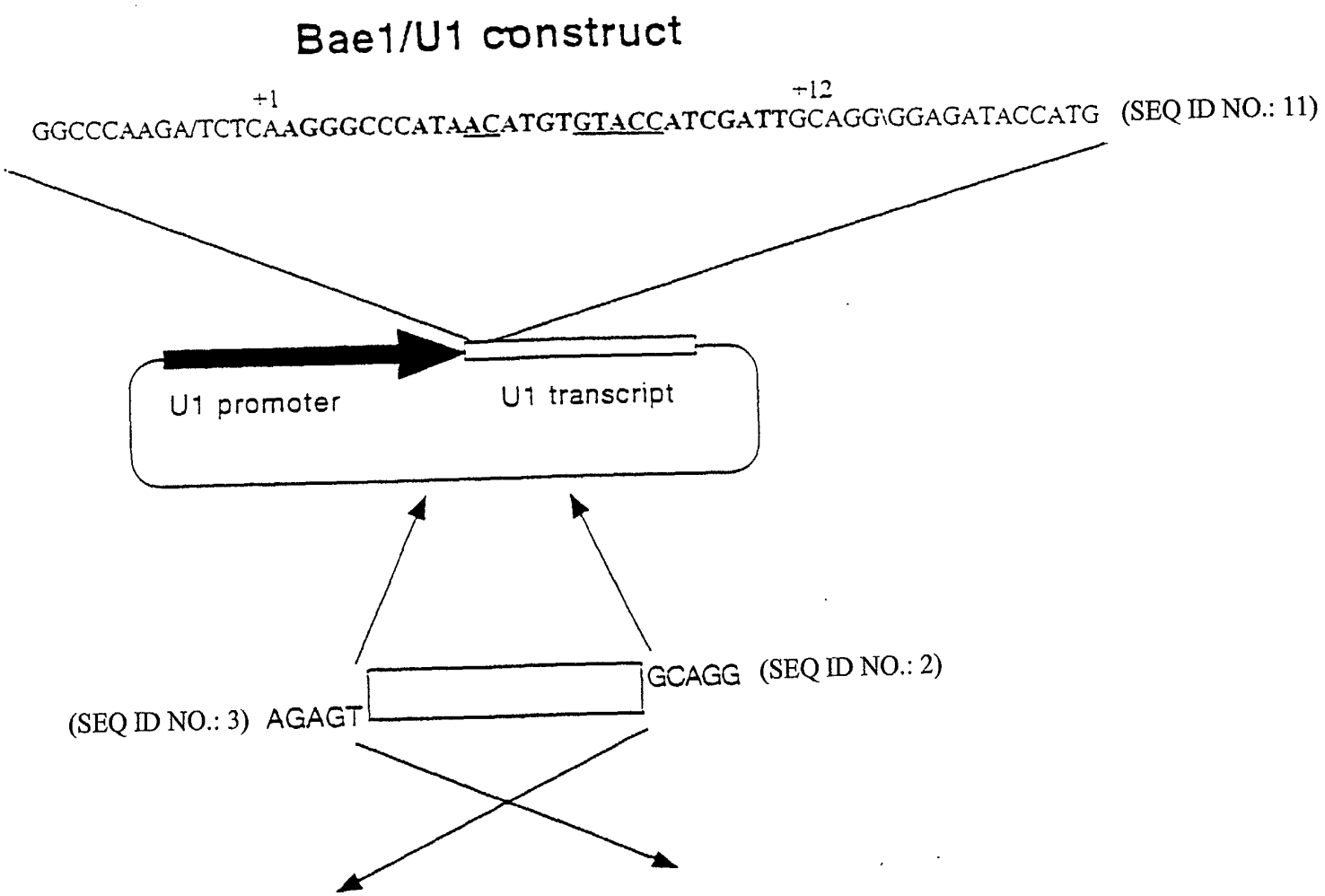

Fig. 4